AD-A093 563 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
THE UBIQUITOUS ROLE OF F'/F IN EFFICIENCY ROBUST ESTIMATION OF --ETC(U)
AUG 80 B L JOINER, D L HALL DAAG29-75-C-0024
UNCLASSIFIED MRC-TSR-2102 NL

1 of 1
AD
A093563

END
DATE
FILMED
2-81
DTIC

MRC Technical Summary Report #2102

THE UBIQUITOUS ROLE OF f'/f IN EFFICIENCY ROBUST ESTIMATION OF LOCATION

Brian L. Joiner and David L. Hall

(1)

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

August 1980

(Received February 22, 1980)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

THE UBIQUITOUS ROLE OF f'/f IN EFFICIENCY ROBUST ESTIMATION OF LOCATION

Brian L. Joiner* and David L. Hall**

Technical Summary Report #2102
August 1980

ABSTRACT

This paper is primarily expository in nature and focuses on the all pervasive importance of f'/f in efficient estimation of location, with primary emphasis on the role of f'/f in robust estimation. Connections between M estimators (maximum likelihood-like), R (rank) estimators and L estimators (linear combinations of order statistics) are discussed and an alternative heuristic explanation of f'/f is given showing why it is an intuitively reasonable quantity on which to base estimation. The asymptotic relative efficiency of each class of estimators is shown to be the square of a correlation coefficient related to f'/f and reasons are given as to why R estimators might often prove to have superior robustness properties relative to L and M estimators.

AMS (MOS) Subject Classifications: 62F35; 62F12; 62F10

Key Words: Fisher information; M estimators; L estimators; R estimators; L-M estimators; Distances between distributions; Asymptotic relative efficiency

Work Unit Number 4 (Statistics and Probability)

*Brian Joiner is Professor and Director of Statistical Laboratory, Department of Statistics, University of Wisconsin-Madison.

**David Hall is Senior Research Scientist, Battelle Northwest Laboratories, Richland, Washington.

Sponsored by the United States Army under Contract Nos. DAAG29-75-C-0024 and DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

There is considerable discussion in statistics as to how one should estimate the location (sometimes called the central tendency) of a distribution. Traditionally the sample mean and its generalization, least squares, have been used, often in conjunction with outlier rejection rules. However, there may be considerable loss in efficiency if the mean or any other preselected estimator is used with data for which it is not appropriate.

This paper provides insight into which characteristics of the parent distribution of the data have a practical impact on the efficiency of the estimator. Three classes of estimators are explored and it is shown that in all three the key quantity is f'/f where f is the density function of the data and f' is its derivative. Correlation coefficients between the f'/f of the hypothesized data and the corresponding quantity g'/g for the actual data, are shown to be directly related to the efficiency of each method of estimation.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

CONTENTS

THE UBIQUITOUS ROLE OF f'/f IN EFFICIENCY ROBUST ESTIMATION OF LOCATION

Brian L. Joiner* and David L. Hall**

1. Introduction

Three major classes of estimators, L, M and R estimators, have been extensively studied in the robustness context but relatively little emphasis has been placed on the similarities and differences among the three classes. An important purpose of this paper is to demonstrate some of their underlying similarities, and in so doing, gain insight as to some of their more important distinctions. The key role of f'/f in these matters is emphasized.

*Brian Joiner is Professor and Director of Statistical Laboratory, Department of Statistics, University of Wisconsin-Madison.

**David Hall is Senior Research Scientist, Battelle Northwest Laboratories, Richland, Washington.

Sponsored by the United States Army under Contract Nos. DAAG29-75-C-0024 and DAAG29-80-C-0041.

2. L, M, and R estimators

In this section we give brief definitions of the three major classes of location estimators and in subsequent sections we describe the relationships among these three classes in large samples.

L-estimators

Of the three classes, the simplest to explain are the L-estimators. An L-estimator of a location parameter λ is of the general form

$$\hat{\lambda} = \sum_{i=1}^{n} a_{i,n} X_{(i)} \qquad (2a)$$

where the $X_{(i)}$ are the ordered observations from a sample of size n and the $a_{i,n}$ are weights to be applied to the various order statistics. A simple example of an L-estimator for a sample or size 4 is

$$\hat{\lambda} = \frac{1}{6} X_{(1)} + \frac{2}{6} X_{(2)} + \frac{2}{6} X_{(3)} + \frac{1}{6} X_{(4)} .$$

In small samples from known distributions the optimal weights for L-estimates are derived from the expectations and variance-covariance matrix of the order statistics by means of the Gauss-Markov theorem. For large samples it is convenient to represent the weights by defining a function $h(u)$ on $(0,1)$ such that $a_{i,n} = h\left(\frac{i}{n+1}\right) / \sum_{i=1}^{n} h\left(\frac{i}{n+1}\right)$. If the data have cdf $F_\lambda(x) = F(x - \lambda)$ with density $f(x - \lambda)$ and if $f'(x - \lambda) \underset{=}{\underline{\text{def}}} \frac{\partial f(x - \lambda)}{\partial x}$, then under regularity conditions it can be shown that the asymptotically most efficient function $h(u)$ for data from F is given by

$$h(u) = g(F^{-1}(u)) , \qquad (2b)$$

where $g(x) = -\frac{d}{dx}(f'/f(x))$, and $F^{-1}(u)$ is the percent point function or inverse cdf of F.

Some examples of optimal h functions are given in Exhibit 2A. The optimal L estimator for Gaussian data is the ordinary sample mean and that for double exponential data is the median. Trimmed means are optimal L-estimators for distributions with Gaussian middles and double exponential tails.

M-estimators

The concept of M-estimators (or maximum-likelihood like estimators) was introduced by

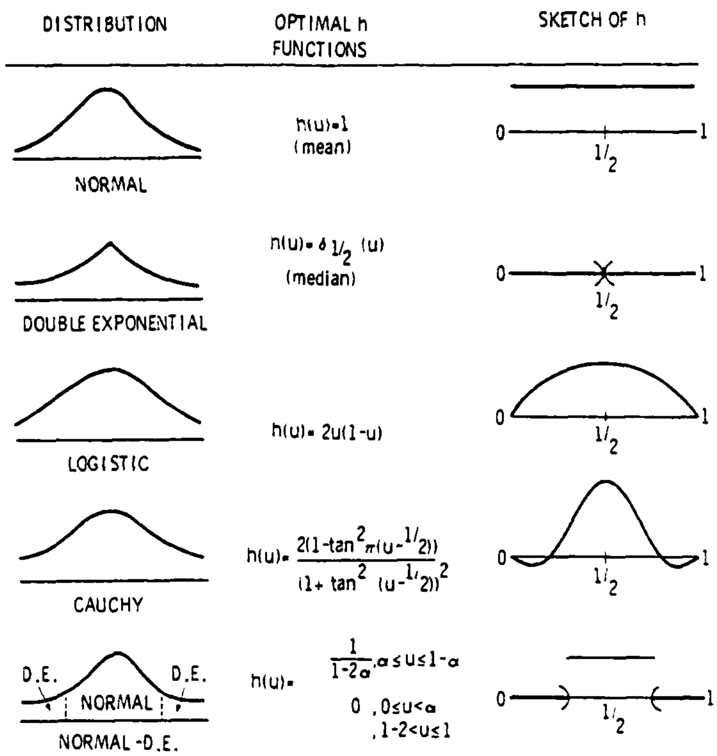

Exhibit 2A

Optimal weight functions for L estimators for several distributions

Huber (1964). In general an M-estimator is defined by a function $\psi(x)$ and the M-estimate of the location parameter λ based upon the data $\{x_i\}$ is given by the value of λ such that $\sum_{i=1}^{n} \psi(x_i - \lambda) = 0$. The asymptotically most efficient M estimator for data from a differentiable density f is the maximum likelihood estimator for which $\psi = -f'/f$.

Some examples of optimal ψ functions are given in Exhibit 2B. Note that for the Gaussian distribution $\psi(z) = z$ and the optimal M estimator is the sample mean, while for the double exponential distribution the best M-estimator is the median. For a distribution with a Gaussian middle and double exponential tails, the maximum likelihood estimator is a metrically trimmed mean in which $\hat{\lambda}$ must be calculated iteratively but winds up being the average of the middle observations after all observations such that $|X_i - \hat{\lambda}| > k$, are trimmed. The points $\lambda \pm k$ are those at which the Gaussian portion of the parent distribution meets the double exponential portions. This metrically trimmed mean is often called a "Huber estimator" since it was found by Huber (1964) to be the minimax estimator for data from a Gaussian distribution with arbitrary symmetric contamination. That is, it is the M-estimator whose worst case variance is minimized over the class of distributions given by $\{(1 - \varepsilon)\Phi + \varepsilon H\}$ where Φ is standard Gaussian and H is symmetric about zero, but otherwise arbitrary.

R-estimators

A class of estimators based on rank tests for symmetry and known as R-estimators was introduced by Hodges and Lehmann (1963). We find these more difficult to explain, but three statements that provide concise and easily understood intuitive definitions for some are:

- Take as an estimator that value of λ for which the rank test scores for the n values $(x_1 - \lambda), (x_2 - \lambda), \ldots, (x_n - \lambda)$ give the best balance relative to the origin (slightly paraphrased from Lehmann, 1975, p. 176);
- An R-estimate of λ is that point of symmetry that is least rejectable by the specified rank test;
- An R-estimate for λ is the midpoint of symmetric confidence intervals for λ based on a specified rank test statistic.

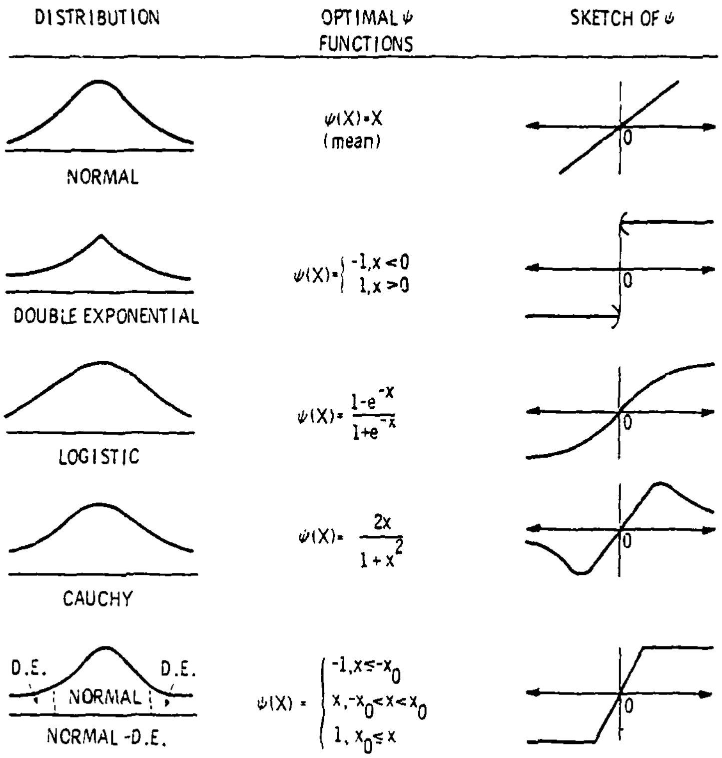

Exhibit 2B

Optimal ψ-functions of M-estimators for several distributions

Proceeding more formally, consider any rank test for symmetry with score function $J(u)$ defined on $0 \leq u \leq 1$ with $J(u) = -J(1 - u)$. Then for any trial value of λ,

(a) compute $\{|X_i - \lambda|\}$,

(b) find ranks of $\{|X_i - \lambda|\}$,

(c) compute $S^+(\lambda) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - \lambda) \cdot J^+\left(\frac{R(|x_i - \lambda|)}{n + 1}\right)$

$$\text{where} \quad J^+(u) = J\left(\tfrac{1}{2} + \tfrac{1}{2} u\right), \quad \text{and} \tag{2c}$$

$$\begin{aligned} \operatorname{sgn}(z) &= +1, \quad \text{if} \quad z > 0\,, \\ &= 0, \quad \text{if} \quad z = 0\,, \\ &= -1, \quad \text{if} \quad z < 0\,. \end{aligned}$$

Then the value of λ such that $S^+(\lambda) = 0$ is the R-estimate corresponding to the score function $J(u)$. If there is no value of λ such that $S^+(\lambda) = 0$, the R estimate is usually defined as the mid-point of the interval between the largest value of λ such that $S^+(\lambda) < 0$ and the smallest value such that $S^+(\lambda) > 0$.

The optimal score function for data from the distribution F is, under some regularity conditions, given by

$$J(u) = -f'/f(F^{-1}(u))\,. \tag{2d}$$

Some examples of optimal score functions and the corresponding rank tests and R-estimators are given in Exhibit 2C. Most R-estimators must be solved iteratively, just like M-estimators. Two important exceptions are the optimal R-estimators corresponding to the double exponential and logistic distributions. For the double exponential distribution the optimal rank test is the sign test and the corresponding optimal R-estimator is the median. For the logistic distribution the Wilcoxon test is optimal as is its counterpart, the Hodges-Lehmann (1963) estimator, defined as the median of the Walsh averages

$$(X_i + X_j)/2 \quad \text{for} \quad i \leq j\,.$$

The optimal R-estimator for the normal distribution must be solved iteratively and correponds to the normal scores test with $J(u) = \Phi^{-1}(u)$, the inverse normal CDF.

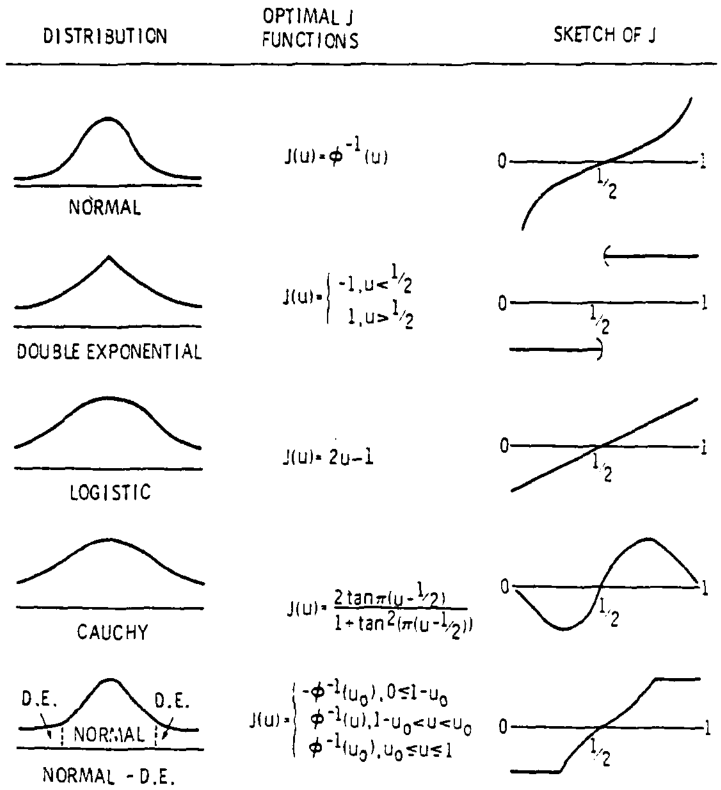

Exhibit 2C

Optimal score functions of R-estimators for several distributions

3. Similarity of optimal L, M and R estimators

In this section we emphasize the very close connections between optimal L, M and R-estimators and show how the L-M-estimators of Rivest (1978) are a natural extension. In all cases we will assume the data come from some symmetric density f known up to a location parameter λ and appropriate regularity conditions on f will be assumed to hold.

Optimal M estimators

As mentioned in the preceding section and as is well known the optimal M-estimator for data from f is given by the maximum likelihood estimator $\hat{\lambda}_M$ defined by

$$\sum f'/f(x_i - \hat{\lambda}_M) = 0 .$$

Optimal R-estimators

The optimal R-estimator given by (2c and 2d) can be shown to be quite similar to the optimal M-estimator, except that the actual deviations $\{(x_i - \lambda)\}$ in the M-estimator are replaced by "predicted deviations"; that is, by the deviations one would predict based on knowledge of f and the ranks of the absolute values of the deviations. Suppose we define

$$\widetilde{(X_i - \lambda)} = \mathrm{sgn}(X_i - \lambda) \cdot F_+^{-1}\left(\frac{R(|X_i - \lambda|)}{n+1}\right) \tag{3a}$$

where F_+ is the cdf of $|x - \lambda|$ for the given f, and F_+^{-1} is its percent point function. Then the optimal rank estimate $\hat{\lambda}_R$ is simply the solution $\hat{\lambda}_R$ of

$$\sum f'/f(\widetilde{X_i - \hat{\lambda}_R}) = 0 .$$

Thus an M-estimator makes use of the <u>actual deviation</u> while a rank estimator must replace the actual deviation by some function of its rank. Since symmetry has been assumed the best value to use is the value one would predict based solely on knowledge of the rank of the absolute value of the deviation. This is an estimate of how big the absolute value of the deviation "should have been" for data from the known f. In formula (3a) the sgn function merely keeps track of which side of $\hat{\lambda}$ the deviation came from, while the other part gives the magnitude of the deviation one would predict.

Optimal L-estimators

For L-estimators a somewhat analogous result can be shown to hold, except here the actual deviation is used just as it was for M-estimators, while f'/f is approximated by a locally linear function. To see this suppose we start from the M-estimator $\sum f'/f(X_{(i)} - \lambda) = 0$ evaluated now for the ordered observations $\{x_{(i)}\}$. Taking some particular deviation, say $(X_{(i)} - \lambda)$, we seek a linear Taylor series approximation for the function $f'/f(X_{(i)} - \lambda)$. Expanding about the value $(X_{(i)} - \lambda)^0$, let

$$f'/f(X_{(i)}-\lambda) \doteq f'/f(X_{(i)}-\lambda)^0 + \left[(X_{(i)}-\lambda) - (X_{(i)}-\lambda)^0\right] \left.\frac{\partial f'/f(X_{(i)}-\lambda)}{\partial(X-\lambda)}\right|_{(X_{(i)}-\lambda)^0} .$$

Now choose

$$(X_{(i)} - \lambda)^0 = F^{-1}\left(\frac{i}{n+1}\right)$$

as the value about which the Taylor series expansion is taken. Then

$$f'/f(X_{(i)}-\lambda) \doteq f'/f\left(F^{-1}\left(\frac{i}{n+1}\right)\right) + (X_{(i)}-\lambda)h\left(\frac{i}{n+1}\right) - F^{-1}\left(\frac{i}{n+1}\right)h\left(\frac{i}{n+1}\right)$$

where $h(u) = \frac{\partial f'/f}{\partial x}(F^{-1}(u))$ as given by (2b). But $\sum f'/f\left(F^{-1}\left(\frac{i}{n+1}\right)\right) = 0$ since this is just the MLE of the center of a distribution symmetric about zero with "data" equal to the symmetric quantiles. Similarly $\sum h\left(\frac{i}{n+1}\right) \cdot F^{-1}\left(\frac{i}{n+1}\right) = 0$ since this is simply an L-estimate of the same center for the same quantile "data". Thus

$$\sum f'/f(X_{(i)} - \lambda) \doteq \sum h\left(\frac{i}{n+1}\right) \cdot (X_{(i)} - \lambda) .$$

When the right hand half of this is set equal to zero, it yields, the L-estimator

$$\hat{\lambda}_L = \sum_{i=1}^{n} h\left(\frac{i}{n+1}\right)X_{(i)} \Big/ \sum_{j=1}^{n} h\left(\frac{j}{n+1}\right) .$$

This process can be seen more clearly in Exhibit 3A. There the curved line is the ψ function that is asymptotically optimal for a Tukey lambda variate with parameter -0.5.*

---

*A Tukey lambda variate z with parameter γ is defined by the equation $z = [U^{\gamma} - (1 - U)^{\gamma}]/\gamma$ where U is uniform on $(0,1)$. See, e.g., Joiner and Rosenblatt (1971).

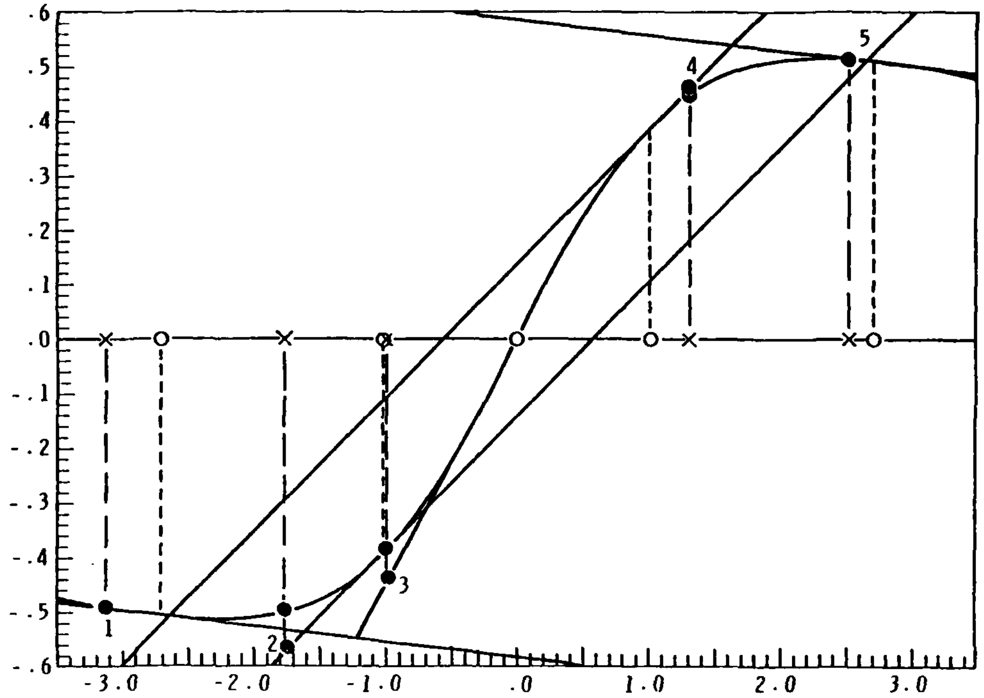

Exhibit 3A

Relationship between asymptotically equivalent M and L estimators

Exhibit 3A Continued

The curved line is a ψ function for an M estimator optimal for a Tukey λ variate (with $\lambda = -.5$). The circles on the horizontal axis are at the points $F^{-1}(\frac{i}{N+1})$ for $N=5$ observations. Tangent lines are drawn to ψ at these points. These tangent lines are the small sample analog of those that define the asymptotically equivalent L estimator. The x's on the horizontal axis denote the observed data. For the M estimator one computes $\Sigma\psi$ at the data, moving the ψ curve left or right until $\Sigma\psi = 0$. The center of the ψ curve then gives $\hat{\theta}_M$. For the L estimator the process is analogous except that one uses the tangent line approximations to ψ. The smallest value of x uses the first tangent line; $x_{(2)}$ uses the second tangent line, and so on. Even in this small sample it is clear that there is little difference between the ψ weights and their L approximations.

The circles on the horizontal axis are at the points $F^{-1}(\frac{i}{n+1})$ for $n = 5$ observations and a trial value of λ. Tangent lines are drawn to ψ at these points. These tangent lines are the small sample analog of those that define the asymptotically equivalent L-estimator. The x's on the horizontal axis denote the observed data. For the M-estimator one computes $\Sigma\psi$ at the data, moving the ψ curve to the left or right until $\Sigma\psi = 0$. The center of the ψ curve then gives the M estimator $\hat{\lambda}_M$.

For the L-estimator the process is conceptually analagous except that the tangent line approximations are used rather than ψ itself. The smallest value of x uses the first tangent line, the second smallest uses the second tangent line and so on. Even in this small sample it is clear that there is little difference between the ψ weights and the weights from the tangent lines. Note also that the tangent lines are not used in the same fashion as the customary piecewise linear approximation. The smallest data value uses the <u>first</u> tangent line no matter how far out (or in) that point might fall.

To sum up, when one knows the parent distribution there is a very close connection among the asymptotically optimal L, M and R-estimators. The M-estimator is the maximum likelihood estimator, with $\psi = f'/f$ and the L and R-estimators are defined by simple approximations. This close connection warrants summarization.

<u>Optimal M-estimator for f is value of λ such that</u>

$$\sum f'/f(X_i - \lambda) = 0 ;$$

<u>Optimal R-estimator for f is value of λ such that</u>

$$\sum f'/f \text{ (predicted value of } (X_i - \lambda)|f \text{ and rank of } |X_i - \lambda|) = 0;$$

<u>Optimal L-estimator for f is value of λ such that</u>

$$\sum [\text{linear approximation of } f'/f](X_{(i)} - \lambda) = 0.$$

<u>L-M estimators</u>

Seeing the intimate connection among these estimators leads one to think of broader classes of estimators that would combine or include these three. The work of Rivest (1978) provides one such class. Rivest has studied a class of L-M estimators defined as the solution of

$$\sum_{n=1}^{n} h\left(\frac{i}{n+1}\right) \cdot \psi(X_{(i)} - \lambda) = 0 \ .$$

These estimators would seem to combine features of both L and M-estimators. As one might conjecture, these estimators turn out to be asymptotically equivalent to maximum likelihood estimators at F if H and ψ are such that

$$h(u) \cdot \frac{\partial \psi}{\partial x} F^{-1}(u) \equiv \frac{\partial f'/f}{\partial x} F^{-1}(u) \ .$$

That is, the product of h, the L component weight function, and the slope of ψ, the M component function, must be identical to the derivative of the maximum likelihood score function. Thus an asymptotically optimal L-M-estimator with h and ψ functions defined by $h(u) = \frac{\partial \psi}{\partial x} F^{-1}(u) = \left[\frac{\partial f'/f}{\partial x} F^{-1}(u)\right]^{1/2}$ would, in some sense, be "half" M and "half" L.

4. Heuristic view of f'/f

In the preceding section we saw that

$$-\frac{f'}{f} = \frac{\partial f(X-\lambda)}{\partial\lambda} \div f(X-\lambda)$$

is the key quantity in efficient estimation of location, be it M, L, or R or even L-M-estimation. This important fact seems not to be widely appreciated even though it is implicit in many sources. In this section we give a heuristic view as to why it is eminently plausible to base estimates on f'/f. This intuitive motivation is intended to compliment that of the likelihood approach.

In the likelihood approach one starts from the fact that the "probability" of the data for any given value of λ is $\prod_{i=1}^{n} f(X_i - \lambda)$. One then finds the value of λ that maximizes the "probability of the data". Taking logarithms and differentiating leads to the familiar $\sum -\frac{f'}{f}(X_i - \lambda) = 0$ as defining the maximum likelihood estimate of λ. Even after seeing this, many of us still have little "feel" as to why f'/f "should be" the defining characteristic.

Here we give an alternative view that seems to be plausible enough even for many students taking their first course in statistics. The exposition is all in the context of estimating the location of a symmetric distribution known up to its point of symmetry, however much is immediately generalizable to broader classes of estimation problems.

The process of estimation can be viewed as essentially the matching of a density with an observed histogram. One might imagine the density function in Exhibit 4A being moved along the horizontal axis until it provides a good match, in some sense, with the observed data. Once the "best" match has been found, the location estimate becomes the center of symmetry of the density function.

The role of f'/f in efficient estimation of location is thus to determine when a good match has been obtained. To see how this is accomplished consider Exhibits 4B and 4C. These present two microscopic views of the interrelationship between the data and the theoretical density function at different regions of the horizontal axis. The amount of information available locally concerning the incremental movement of the density relative to the data is quite different at the two sites. In Exhibit 4B the local portion of the

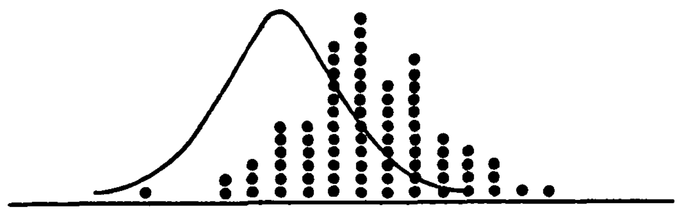

Exhibit 4A

Heuristic view of estimation of location. The density is moved alone until it "matches" the data as well as possible. Matches are determined by the balancing of stresses due to the relative steepness of the density function at various points along the horizontal axis, i.e. by making $\sum -\frac{f'}{f} = 0$. The center of symmetry of this density is then the estimate of location for the set of data.

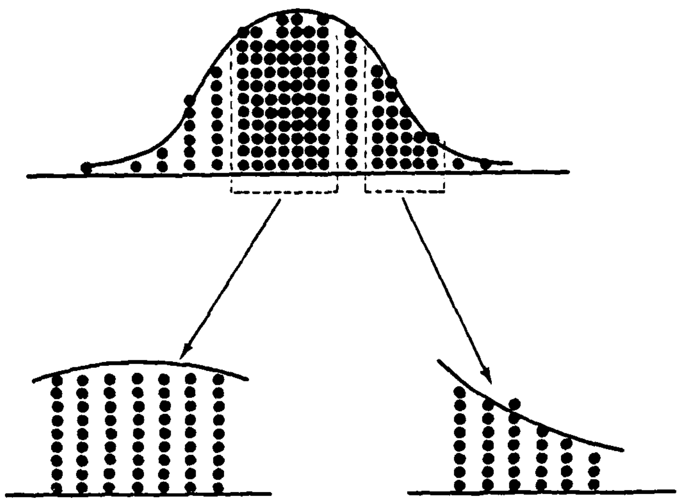

Exhibit 4B

Close-up of portion of density-data where $\frac{f'}{f} \doteq 0$.

Exhibit 4C

Close-up of portion of density-data where $\frac{f'}{f}$ is large

Close-up views of two portions of the density function and the data it is seeking to fit; one view where $\frac{f'}{f}$ is essentially zero, and thus where there is little local information about whether the density is too far left or right relative to the data; and one where $\frac{f'}{f}$ is large and much can be learned locally about appropriate placement of the density function relative to the data.

density would fit the local data almost as well if the density were moved slightly to either side. Thus there is very little local information that could be used to determined the value of λ.

In Exhibit 4C the situation is quite different. In this case any movement of the density would markedly decrease the quality of the match between the density and the data, and thus there is much relevant local information concerning estimation of location.

The main difference between Exhibits 4B and 4C is in the relative steepness of the local part of the density function, i.e. in the magnitude of f'/f. However, even after considering this, one might still question why f'/f rather than, say f' alone is the key characteristic in estimation of location. While the answer is not obvious it does seem heuristically reasonable that the height of a histogram should also be relevant since a given amount of tilt at the top of a tall histogram may well be less informative than the same amount in a very short one.

Thus, seen from this view, the role of f'/f is to measure the relative steepness of the density function and to express the amount of resistance the data exerts to having λ moved away from it. Once λ is near the correct value, there will exist stresses from f'/f on both sides of λ. For any given sample, the value of λ that balances these stresses will be the location estimate for that particular sample.

It is interesting to review these "stress functions" for several well known distributions. Exhibit 2B presents $-f'/f$ for several distributions. When considered from the above viewpoint, the fact that for the <u>Gaussian</u> distribution $-f'/f$ is <u>exactly proportional</u> to the size of the deviation is quite remarkable. The tails of a Gaussian distribution thus get ever increasingly steep as one moves further away from its center in <u>direct proportion</u> to the distance. Hence for a Gaussian distribution the amount of resistance exerted to movement of λ away from an observation increases in direct proportion to the magnitude of the deviation. This, of course, is the cause of the well known phenomenon that outlying observations have great effect on location estimation for M estimators based on the Gaussian assumption.

The double exponential distribution, on the other hand, has tails of constant relative steepness. Thus the stress exerted on λ does not depend at all on how far out in the tails an observation is. All that matters is what side of λ the observation is on.

The logistic distribution differs from the double exponential in its central portion but, as can be seen from Exhibit 2B, its tails are asymptotically equivalent to those of the double exponential. Thus, once λ is a substantial distance from an observation, moving it an arbitrary amount further away makes virtually no difference in the amount of force exerted on λ by that observation under logistic assumptions.

The Cauchy distribution is different yet: observations at an intermediate distance from λ exert the most force while those further out exert almost none. Under Cauchy assumptions, once an observation is far enough away from λ, not even the side it is on matters much. Cauchy tails are asymptotically flat, like a uniform distribution, and thus contain essentially no information on location. The greatest information concerning location in a Cauchy distribution is in the "shoulders" of the density which fall off rather sharply.

The uniform distribution itself represents a limiting case in another direction. Here, f'/f shows that the middle part of the data contain no information on location while the endpoints contain "infinite" information, thus λ must exactly balance the two endpoints, leading to the well known result that the midrange is the MLE and asymptotically most efficient L estimator for the uniform distribution.

The point of view discussed above is naturally quite closely connected to the idea of the influence curve introduced by Hampel (1968). Hampel noted that if a derivative of the functional defining an estimator was taken, the resulting function $\Omega(X,F)$ could be interpreted as representing how much effect an observation at x would have on the estimator with data from F.

A Mechanical Analogy

These observations lead us to propose an alternative mechanical view of estimation of location. The conventional view is that of finding the balance point of a scale in which blocks of equal mass have been placed at each data point. This analogy is exact for the

Gaussian case and can be modified to work in other cases by using an analogy with weighted estimation (c.f., Andrews, 1974).

An alternative view that seems to provide different insight is that of the same balance, but with blocks of mass proportional to $|\frac{f'}{f}(X_i - \lambda)|$ being placed at ± 1, depending on the sign of $(X_i - \lambda)$. The masses represent the amount of stress or force being exerted by the various data points as a function of the relative steepness of f'/f at that distance from λ. For the double exponential, all masses would be of the same size so that λ only seeks to have an equal <u>number</u> of masses on each side. For the Cauchy, the masses would first increase then decrease in size.

<u>Other Uses of f'/f</u>

That f'/f plays a key role in estimation is of course not new. It is explicitly the central quantity in maximum likelihood estimation and is at least implicit in L and R-estimation. Fisher information being equal to $E(f'/f)^2$ is thus a measure of the "amount" of f'/f. The Cramer-Rao lower bound for the variance of an estimator of location being equal to $\frac{1}{E(f'/f)^2}$ has an analogous interpretation.

Stein (1956), Stone (1975) and others have shown that at least asymptotically for symmetric distributions it is possible to estimate f'/f from the data and thus gain full asymptotic efficiency for data from any symmetric distribution, subject to mild regularity constraints. In fact Stone's results can be said to be promising even for the location problem even in samples as small as 40.

<u>Huber's Minimax Result</u>

A less obvious situation in which f'/f appears to be key is associated with Huber's (1964) minimax estimator for a type of contaminated data. Huber proposed the following problem: suppose F is a symmetric distribution known up to a location parameter and H is any other distribution symmetric about the same point. Then consider the class of contaminated distributions $\{(1 - \varepsilon)F + \varepsilon H\}$ where ε is fixed. He asked, which fixed estimator has the best worst-case variance with respect to this class of distributions?

Huber showed that if F had a strongly unimodal density, i.e. if $-f'/f$ were monotonic, then under mild regularity conditions the solution was given by an M estimator

of the form

$$\sum_{i=1}^{N} \psi(X_i - \lambda) = 0 ,$$

where

$$\psi(x) = -\frac{f'}{f}(x), \quad \text{for} \quad |x| \leq c ,$$

$$= -\text{sgn}(x)\left(\frac{f'}{f}\right)(c), \quad \text{for} \quad |x| > c ,$$

and where c is determined by ε. Huber interpreted this result as a sort of "fattening up" of F's tails. However, it seems more pertinent to view it as the removal of the most informative part of f'/f. For example, if F is Gaussian then $-f'/f$ is a straight line with positive slope. The minimax estimator for the above class is known as a "Huber" and has the ψ function shown at the bottom of Exhibit 2B.

Thus nature's best strategy is to take as H that distribution which places all of its mass in the portions of F that have the greatest relative steepness in such a way as to make *those* portions of the resulting density exponential. Hence the worst possible Huber-type contaminated normal has a Gaussian middle and double exponential tails, and has as its maximum likelihood estimator the M estimator defined above.

A Conjecture

Huber's proof makes critical use of the assumed strong unimodality of F and thus does not apply when F is a distribution like the Cauchy. However, we conjecture that Huber's result holds in a broader class of distributions in the sense that the Huber-type minimax estimators for any distribution will, under reasonable regularity conditions, be of the form

$$\psi(x) = -\frac{f'}{f}(x), \qquad x \in A ,$$

$$= -\text{sgn}(x) \cdot K, \qquad x \notin A ,$$

where

$$A = \{x : |\frac{f'}{f}(x)| \leq K\} .$$

5. f'/f and relative efficiency of estimation

In Section 3 we observed that f'/f is the key quantity in defining a fully efficient estimator of location. In this section we investigate the problem of relative efficiency of estimation, where one uses an estimator optimal for data from F, but applies it to data which actually came from some other distribution G. We show that in such cases the asymptotic relative efficiency (ARE) of L, M and R-estimators are all determined by correlation coefficients between f'/f and g'/g. The difference in efficiencies among L, M and R-estimators is shown to be a matter of the "data" at which f'/f is evaluated. This provides us with insight as to differences among the three classes.

Efficiencies as Squared Correlations

Correlation coefficients occur frequently in efficiency calculations. Cramer (1945) showed that if T_1 was an efficient estimator and T_2 was a regular unbiased estimator, then the square of the correlation coefficient between the estimators gave the efficiency of T_2. Noether (1955), Hajek (1962), and van Eeden (1963) extended this result and showed in different situations that the Pitman efficiency of certain tests was given by the square of the correlation coefficient between the test statistics. In the context of rank tests this correlation between the rank statistics reduces to the correlation between the asymptotic score functions corresponding to the tests (Hajek, 1962). (Note that it is often much easier to compute the correlation between score functions than it is among the estimators which they define.)

R-Estimators

First we show that the ARE of a rank estimator corresponding to an arbitrary score function is given by the square of a type of correlation coefficient. Hajek (1962) showed that when the two sample rank test based on the score function $J(u)$ is applied to data from a distribution F, the ARE of the test based on J with respect to the asymptotically most powerful rank test (amprt) for the distribution F is given by the square of a correlation coefficient, namely

$$\rho^2(J(u), -f'/f(F^{-1}(u))) = \frac{\left[\int_0^1 J(u)(-f'/f(F^{-1}(u))du\right]^2}{\int_0^1 J^2(u)du \int_0^1 [-f'/f(F^{-1}(u))]^2 du} . \qquad (5.1)$$

van Eeden (1963) proved a similar result for the one sample test for symmetry. Hajek's result (5.1) is also true for the corresponding rank estimators, as will now be shown directly.

We assume throughout this section that all distributions considered are symmetric and unimodal with finite Fisher information and a differentiable density.

Theorem 5.1: If the R-estimator with score function $J(u)$ is used on data from a distribution F, the ARE of the estimator based on $J(u)$ relative to the R-estimator corresponding to the amprt for F is given by (5.1).

Proof: We can assume without loss of generality that F has been scaled so that its Fisher information

$$I(F) = \int_0^1 [-f'/f(F^{-1}(u))] du$$

is unity. The asymptotic variance of the rank estimator with score function $J(u)$ on the distribution F is given by

$$\sigma_R^2 = \frac{\int_{-\infty}^{\infty} J^2(F(x))f(x)dx}{\left[\int_{-\infty}^{\infty} J'(F(x))f^2(x)dx\right]^2} .$$

Now since $I(F) = 1$, the ARE of this rank estimator with respect to the efficient rank estimator for F is just the reciprocal of σ_R^2. Thus

$$\mathrm{ARE}_R(J|F) = \frac{1}{\sigma_R^2} = \frac{\left[\int_{-\infty}^{\infty} J'(F(x))f^2(x)dx\right]^2}{\int_{-\infty}^{\infty} J^2(F(x))f(x)dx} = \frac{\left[\int_0^1 J'(u)f(F^{-1}(u))du\right]^2}{\int_0^1 J^2(u)du} ,$$

which after integrating the numerator by parts and recalling that $I(F) = 1$ gives

$$\mathrm{ARE}_R(J|F) = \frac{\left[\int_0^1 J(u)f'/f(F^{-1}(u))du\right]^2}{\int_0^1 J^2(u)du \cdot \int_0^1 [-f'/f(F^{-1}(u))]^2du} ,$$

which is the desired result.

In the above, note that if J is the score function for the amprt corresponding to some distribution G then the expression for the ARE becomes

$$\mathrm{ARE}_R(G|F) = \rho^2(-f'/f(F^{-1}(u)), -g'/g(G^{-1}(u)))$$

$$= \frac{\left[\int_0^1 g'/g(G^{-1}(u))f'/f(F^{-1}(u))du\right]^2}{\int_0^1 [f'/f(F^{-1}(u))]^2du \int_0^1 [g'/g(G^{-1}(u))]^2du} . \qquad (5.2)$$

Note that this is the square of the correlation coefficient between f'/f and g'/g with each being evaluated at its own data. Also note the reflexivity of the ARE for rank estimators. That is (5.2) represents the ARE of the rank estimator with score function $-g'/g(G^{-1}(u))$ on the data from F as well as the ARE of the rank estimator with score function $-f'/f(F^{-1}(u))$ on data from G. As an example, the best rank estimator for Gaussian data, the normal scores estimator, has the same ARE on logistic data, 0.95, as the best rank estimator for logistic data, the Hodges-Lehmann estimator has on Gaussian data.

The R efficiencies for a number of other pairs of distributions are computed in Hall and Joiner (1980b).

Correlations, Angles and Efficiencies of R-estimators

For R estimators Gastwirth (1966) has noted that the score function $J(u)$ of a rank estimator may be thought of as an infinite dimensional vector. The score function for the efficient R estimator for F is given by $-f'/f\ F^{-1}(u)$, which may thus also be thought of as an infinite dimensional vector. The square of the cosine of the angle between these two vectors is the ARE of J applied to data from F. This relationship between the ARE of R estimators and the angles between score functions is further developed in Hall and Joiner (1980c).

M Estimators

For M estimators a similar but different result is attained:

Theorem 5.2: The ARE of the M estimator defined by the square integrable function $\psi(x)$ on data from F with respect to the efficient M estimator for F is given by

$$\mathrm{ARE}_M(\psi|F) = \rho^2(\psi(F^{-1}(u)),\ -f'/f(F^{-1}(u)))$$

$$= \frac{\left[\int_0^1 \psi(F^{-1}(u))(-f'/f(F^{-1}(u)))du\right]^2}{\int_0^1 \psi^2(F^{-1}(u))du \int_0^1 (-f'/f(F^{-1}(u)))^2 du} .$$

Proof: The ARE of the M-estimator corresponding to $\psi(x)$ with respect to the efficient M estimator for F, which corresponds to $-f'/f(x)$, is

$$\mathrm{ARE}_M(\psi|F) = \frac{1}{\sigma_M^2 \cdot I(F)}$$

$$= \frac{\left[\int_{-\infty}^{\infty} \psi'(x)F(x)dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x)f(x)dx \int_0^1 [f'/f\ F^{-1}(u)]^2 du}$$

$$= \frac{\left[\int_{-\infty}^{\infty} \psi(x)f'(x)dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x)f(x)dx \int_0^1 [f'/f\ F^{-1}(u)]^2 du}$$

$$= \frac{\left[\int_{-\infty}^{\infty} \psi(x) \frac{f'(x)}{f(x)} f(x)dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x)f(x)dx \int_0^1 [f'/f\ F^{-1}(u)]^2 du}$$

$$= \frac{\left[\int_0^1 \psi(F^{-1}(u))f'/f(F^{-1}(u))du\right]^2}{\int_0^1 \psi^2(F^{-1}(u))du \int_0^1 [f'/f(F^{-1}(u)]^2 du} , \qquad (5.3)$$

which is the desired result. Hence the ARE of an M estimator is also given by the square of a correlation coefficient.

From the above theorem the ARE of the optimal M estimator (i.e. the maximum likelihood estimator) for G when applied to data from F is given by the squared correlation coefficient

$$ARE_M(G|F) = \rho^2(-g'/g(F^{-1}(u)), -f'/f(F^{-1}(u)) .$$

As with R-estimators, the ARE is the squared correlation coefficient between f'/f and g'/g. The big difference here is that f'/f and g'/g are both evaluated at the actual data.

The Role of Scale in M Estimation

In M estimation the scale of the data makes an important difference in estimation of location. For example, the ARE of the M estimator $\psi(x - \lambda)$ when applied to data from $F\left(\frac{x - \lambda}{\sigma}\right)$ depends very much on the value of σ. In R and L estimation, the value of σ is not a factor in determining ARE. This independence of scale in L and R estimation is a convenience, both practically and theoretically.

An illuminating example of M scale dependence is afforded by the family of scaled t distributions. For this family the optimal ψ functions all have identical shape:

$$\psi_\nu\left(\frac{x - \lambda}{\sigma}\right) = (\nu + 1) \cdot \frac{u}{[1 + u^2]} ,$$

where $u = \left(\frac{x - \lambda}{\nu\sigma}\right)$. The roles of σ and ν are thus totally confounded in M estimation

for the t family. One can achieve, for example, 100% efficiency with the ψ for, any t, say the Cauchy, on data from any other t just by using a "wrong" value of σ.

Lack of Reflexivity of M Efficiencies

The reflexivity of efficiencies that R estimators possess is not attained by M estimators. That is, the $ARE_M(F|G)$ is in general different from $ARE_M(G|F)$. The amount of difference depends in general on the scaling of the distributions. For example, in Exhibit 5A we see that

ARE_M(Cauchy|logistic) is not equal to

ARE_M(logistic|Cauchy) for any of the four choices of scale considered.

For R-estimators,

ARE_R(logistic|Cauchy) = ARE_R(Cauchy|logistic) = $6/\pi^2$ = 60.79%.

An even more extreme example of lack of reflexivity is provided by the Gaussian and Cauchy, where

ARE_M(Gaussian|Cauchy) = 0, for all choices of scale.

On the other hand ARE_M(Cauchy|Gaussian) is positive for all choices of scale, and is 57% when the two distributions are expressed in their standard form. The ARE_R is 43%, either way, no matter what scale is used.

L Estimators

The ARE of L estimators is also given by a squared correlation coefficient.

Theorem 5.3: The ARE of the L estimator with weight function $h(u)$, where $h(1-u) = h(u)$ and $\int_0^1 h(u)du = 1$, on data from F with respect to the efficient L estimator for F is given by

$$ARE_L(h|F) = \rho^2[A((h,F)(u)), -f'/f(F^{-1}(u))]$$

where

$$A(h,F)(u) = \int_{1/2}^{u} \frac{h(t)}{f(F^{-1}(t))}\,dt = \int_{1/2}^{u} h(t)d(F^{-1}(t)) \; . \qquad (5.4)$$

Exhibit 5A

Illustration of strong dependence of asymptotic relative efficiency of M estimators on choice of scaling function. All but one of the scaling functions are analogous to the median absolute deviation in that they are a percentile of the $\{|y_i-\hat{y}_i|\}$. For example, MAD = $S_{.50}$ = [0.50 quantile of $\{|y_i-\hat{y}_i|\}$]. Efficiencies for the same ψ function on the same data, range from, e.g., 56% to 75% depending on the choice of scaling function.

(a)

Maximum likelihood estimator for Cauchy applied to logistic data, and vice versa. Note that for the scaling function $S_{.1}$, both M estimators have efficiencies higher than the 60.8% of their rank counterparts.

Efficiencies (in %)

Estimator MLE for	Applied to data from	Scaling function *			
		$S_{.1}$	$S_{.50}$	$S_{.67}$	$(\text{Info})^{-\frac{1}{2}}$
Cauchy	logistic	81.6	77.0	71.2	80.6
logistic	Cauchy	61.4	57.2	52.2	60.4

*Each S was multipled by a k such that 100% efficiency was attached by the MLE on its own data. Thus, for example the logistic estimator had $k_p = [S_p \text{ (on logistic data)}]^{-1}$.

(b)

Tukey bisquare applied to Student t data. The tuning constant k in the bisquare was, in each case, selected to produce 95% efficiency on normal data.

	Efficiencies (in %)		
	Scaling function		
Data from	$S_{.25}$	$S_{.50}$	$S_{.75}$
k	15.268	7.213	4.229
Cauchy	75.3	70.2	56.3
t with $\nu = 2$	90.8	89.7	86.1
$\nu = 3$	95.7	95.3	94.1
$\nu = 5$	98.4	98.4	98.3
$\nu = 10$	98.7	98.8	99.0
$\nu = 30$	97.2	97.3	97.4
normal	95.0	95.0	95.0

The results in this Exhibit are due to Lane Bishop.

Proof: Since $A(h,F)(u) = \int_{1/2}^{u} \frac{h(t)}{f(F^{-1}(t))} dt$, $h(u) = f(F^{-1}(u))A'(h,F)(u)$, where $A'(h,F)(u) = \frac{d}{du} A(h,F)(u)$. Thus, since $\int_0^1 h(u)du = 1$,

$$\int_0^1 f(F^{-1}(u))A'(h,F)(u)du = 1 .$$

Integrating by parts gives that

$$-\int_0^1 A(h,F)(u)d(f(F^{-1}(u))) = 1 .$$

Now assume without loss of generality that $I(F) = 1$ and recall (see, e.g., Huber, 1972) that the asymptotic variance of an L estimator is given by $\sigma_L^2 = \int_0^1 A^2(h,F)(u)du$. Then the ARE of the L estimator $h(u)$ on data from F is given by

$$ARE_L(h|F) = \frac{1}{\sigma_L^2}$$

$$= \frac{[\int_0^1 A(h,F)(u)d(f(F^{-1}(u)))]^2}{\int_0^1 A^2(h,F)(u)du}$$

$$= \frac{\int_0^1 A(h,F)(u)f'/f(F^{-1}(u))du]^2}{\int_0^1 A^2(h,F)(u)du \int_0^1 [f'/f(F^{-1}(u))]^2 du}$$

which was to be shown.

When $h(u) = (-g'/g)'(G^{-1}(u))$, so that it is the optimal weight function for the distribution G, the ARE of it on data from F is:

$$ARE_L(G|F) = \rho^2[A(-(g'/g)'(G^{-1}(u)),F)(u), -f'/f(F^{-1}(u))] .$$

If we let h_f denote the efficient weight function for F, we have

$$A(h_f,F)(u) = \int_{1/2}^{u} \frac{h_f(t)}{f(F^{-1}(t))}\,dt$$

$$= -\int_{1/2}^{u} \frac{d}{dt}\,(-f'/f(F^{-1}(t)))\,dt$$

$$= -f'/f(F^{-1}(u)) \;.$$

If we now let h_g denote the efficient weight function for G, equation (5.4) can be expressed more symmetrically as

$$ARE_L(G|F) = \rho^2(A(h_g,F)(u),A(h_f,F)(u)) \;.$$

Like M-estimators, L-estimators do not have the reflexivity of efficiency possessed by R-estimators. For example, the mean, which is the efficient L-estimator (as well as M-estimator) for the Gaussian distribution, has an ARE of 50% when used on double exponential data; while the median, which is the efficient L estimator (as well as R-estimator) for the double exponential distribuiton, has an ARE of 64% when used on Gaussian data. The *corresponding* R-estimators have ARE equal to 64% in both cases.

Relationships

It is useful to emphasis the similarities and differences among the correlation coefficients for the three classes of estimators. Repeating the ARE formulas derived above we have:

$$ARE_R(G|F) = \rho^2[-g'/g(G^{-1}(u)),\ -f'/f(F^{-1}(u))] \;,$$

$$ARE_M(G|F) = \rho^2[-g'/g(F^{-1}(u)),\ -f'/f(F^{-1}(u))], \text{ and}$$

$$ARE_L(G|F) = \rho^2[A(h_g,F)(u),\ -f'/f(F^{-1}(u))] \;.$$

Note that in R estimation asymptotic relative efficiency is determined by how well f'/f and g'/g correlate when each is evaluated at its own data. In contrast, for M estimation, asymptotic relative efficiency is determined by how well f'/f and g'/g correlate when both are evaluated at the actual data, $F^{-1}(u)$. For L estimation the situation is different still. In this case linear approximations to f'/f and g'/g are

evaluated at their own data and then smoothed by the distribution of the actual data. Asymptotic relative efficiency for L depends upon how well thse two smoothed versions correlated.

A Further Connection

An obvious similarity among the correlation coefficients is the presence of $-f'/f(F^{-1}(u))$ in each of them when the data are from the distribution F. Thus the correlations may all be interpreted as being with the optimal rank score function for the actual data. We can couple this interpretation with the result of Hajek (1962) that for any rank score function one can find a corresponding distribution whose amprt has that score function. With appropriate conditions on ψ and h it would thus be possible to find distributions $G_{\psi,F}$ and $G_{h,F}$ with $I(G_{\psi,F}) = I(G_{h,F}) = 1$ such that

$$\frac{-g'_{\psi,F}}{g_{\psi,F}}(G^{-1}_{\psi,F}(u)) = \frac{\psi(F^{-1}(u))}{\int_0^1 \psi^2(F^{-1}(u))du},$$

and

$$\frac{-g'_{h,F}}{g_{h,F}}(G^{-1}_{h,F}(u)) = \frac{A(h,F)(u)}{\int_0^1 A^2(h,F)(u)du}.$$

That is, the functions

$$\frac{\psi(F^{-1}(u))}{\int_0^1 \psi^2(F^{-1}(u))du}$$

and

$$\frac{A(h,F)(u)}{\int_0^1 A^2(h,F)(u)du}$$

would correspond to the score functions of the amprt's for the distributions $G_{\psi,F}$ and $G_{h,F}$, respectively. Thus the score functions for rank estimators and their correlations contain information concerning not only the efficiencies of rank estimators but also implicitly the efficiencies of M and L estimators. This approach might conceivably be useful in extending the interpretation of some of the results presented in Joiner and Hall (1979).

6. Conclusions

This paper has emphasized the important role played by f'/f in determining the efficiency of all three major classes of location estimators: L, M, and R. In all three, f'/f is used to define the asymptotically efficient estimator and a heuristic view is given as to why f'/f might be an intuitively reasonable quantity on which to base location estimation. The asymptotic relative efficiency of each of the three classes of estimators is seen to depend upon the degree of agreement between f'/f of the hypothesized distribution and the corresponding quantity g'/g for the distribution which actually generated the data.

The insight gained in this paper is used, in several companion papers, to develop other results useful in robust estimation. In Hall and Joiner (1980b) a number of numerical and analytical results are given for the asymptotic relative efficiencies of R estimators optimal for some distribution F when applied to data from some other distribution G. Then in Hall and Joiner (1980c) the R efficiencies are used to develop several useful low dimensional representations of the space of distributions. Underway is a quantative comparison of the relative efficiencies among the three classes of estimators (Joiner and Hall, 1980d). This comparison was prompted by the relationship noted here that in the correlation coefficient that determines R efficiency, f'/f and g'/g are both evaluated at their own data, while for M and L estimation the hypothesized estimator is, at least in part, evaluated at the actual data. This suggests the possibility of some general efficiency robustness for R estimators. However, Bishop's result (cited in Exhibit 5A) shows that there exist pairs of distributions and choices of scaling functions for which both M estimators (i.e., for F on G data and for G on F data) have better ARE than their R counterparts.

In still another related paper (Joiner, Hall and Bishop, 1980e) the close relationship between the defining equations for L, M and R estimators is used to extend these results to the general linear model. This extension is equivalent to what Bickel (1976) has called "pseudo observations" in the context of M estimators.

7. Acknowledgements

We are indebted to George Tiao, Peter Bickel and especially to Lane Bishop for insightful comments. We are also much appreciative of the typing skills of Debbie Dickson, Mary Hall and Sally Ross.

8. References

Andrews, David (1974). "A robust method for multiple linear regression", Technometrics 16, 523-531.

Bickel, Peter J. (1976). "Reply to Discussion of 'Another look at robustness: A review of reviews and some new developments'", Scandanavian J. Statistics 3, 166-168.

Cramer, H. (1956). Mathematical Methods of Statistics. Princeton, N.J.: Princeton University Press.

Gastwirth, J. (1966). "On robust procedures", J. Amer. Statist. Assoc. 61, 929-948.

Hajek, J. (1962). "Asymptotically most powerful rank order tests", Ann. Math. Statist. 33, 1124-1147.

Hall, David L. and Brian L. Joiner (1980b). "Asymptotic relative efficiency of R-estimators: numerical and analytic results", (unpublished manuscript).

Hall, David L. and Brian L. Joiner (1980c). "Representations of the space of distributions useful in robust estimation of location", (unpublished manuscript).

Hampel, F. R. (1968). Contributions to the Theory of Robust Estimation. Ph.D. thesis, University of California, Berkeley.

Hodges, J. L. and E. L. Lehmann (1963). "Estimates of location based on rank tests", Ann. Math. Statist. 34, 598-611.

Huber, P. J. (1964). "Robust estimation of location parameter", Ann. Math. Statist. 35, 73-101.

Joiner, Brian L. and David L. Hall (1980d). "A quantative comparison of the relative efficiency of L, M, and R estimators", (manuscript in preparation).

Joiner, Brian L., David L. Hall and Lane Bishop (1980e). "Robust regression using M, R and L estimators", (manuscript in preparation).

Joiner, Brian L. and Joan R. Rosenblatt (1971). "Some properties of the range in samples from Tukey's symmetric lambda distributions", J. Amer. Statist. Assoc. 66, 394-399.

Lehmann, E. L. (1975). Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day.

Noether, G. E. (1955). "On a theorem of Pitman", Ann. Math. Statist. 26, 64-68.

Rivest, L. P. (1978). On a Class of Estimators of the Location Parameter Based on a Weighted Sum of the Observations. Ph.D. thesis, McGill University.

Stein, Charles (1956). "Efficient nonparametric testing and estimation", Proc. Third Berkeley Symp. on Math. Statist. and Probability, U. of California Press.

Stone, C. (1975). "Adaptive maximum likelihood estimators of a location parameter", Ann. Statist. 3, 267-284.

van Eeden, C. (1963). "The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics", Ann. Math. Statist. 34, 1442-1451.

BLJ/DLH/scr

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER [illegible]	2. GOVT ACCESSION NO. AD A093563	3. RECIPIENT'S CATALOG NUMBER [illegible]
4. TITLE (and Subtitle) THE UBIQUITOUS [illegible] f'/f IN EFFICIENCY ROBUST ESTIMATION OF LOCATION		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Brian L. Joiner [redacted] David L. Hall		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024 DAAG29-[illegible]
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit # 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE August [illegible]
		13. NUMBER OF PAGES [illegible]
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12		15. SECURITY CLASS (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fisher information
M estimators
L estimators
R estimators
L-M estimators
Distances between distributions
Asymptotic relative efficiency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper is primarily expository in nature and focuses on the all pervasive importance of f'/f in efficient estimation of location, with primary emphasis on the role of f'/f in robust estimation. Connections between M estimators (maximum likelihood-like), R (rank) estimators and L estimators (linear combinations of order statistics) are discussed and an alternative heuristic explanation of f'/f is given showing why it is an intuitively reasonable quantity on which to base estimation. The asymptotic relative efficiency

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT - Cont'd.

of each class of estimators is shown to be the square of a correlation coefficient related to f'/f and reasons are given as to why R estimators might often prove to have superior robustness properties relative to L and M estimators.

LME